

Concomitants and majorization bounds for bivariate distribution function

Ismihan Bairamov

Department of Mathematics, Izmir University of Economics
Izmir, Turkey. E-mail: ismihan.bayramoglu@ieu.edu.tr

September 8, 2011

Abstract

Let (X, Y) be a random vector with distribution function $F(x, y)$, and $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ are independent copies of (X, Y) . Let $X_{i:n}$ be the i th order statistics constructed from the sample X_1, X_2, \dots, X_n of the first coordinate of the bivariate sample and $Y_{[i:n]}$ be the concomitant of $X_{i:n}$. Denote $F_{i:n}(x, y) = P\{X_{i:n} \leq x, Y_{[i:n]} \leq y\}$. Using majorization theory we write upper and lower bounds for F expressed in terms of mixtures of joint distributions of order statistics and their concomitants, i.e. $\sum_{i=1}^n p_i F_{i:n}(x, y)$ and $\sum_{i=1}^n p_i F_{n-i+1:n}(x, y)$. It is shown that these bounds converge to F for a particular sequence $(p_1(m), p_2(m), \dots, p_n(m))$, $m = 1, 2, \dots$ as $m \rightarrow \infty$.

1 Introduction

Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be independent and identically distributed (iid) random vectors with joint distribution function (cdf) $F(x, y)$ and $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ be the order statistics of the sample of first coordinate X_1, X_2, \dots, X_n . Denote the Y -variate associated with $X_{i:n}$ by $Y_{[i:n]}$, $i = 1, 2, \dots, n$, i.e. $Y_{[i:n]} = Y_k$ iff $X_{i:n} = X_k$. The random variables $Y_{[1:n]}, Y_{[2:n]}, \dots, Y_{[n:n]}$ are called concomitants of order statistics $X_{1:n}, X_{2:n}, \dots, X_{n:n}$. The theory of order statistics is well documented in David (1981), David and Nagaraja (2003), Arnold et al. (1992). The concomitants of order statistics are described in David (1973), Bhattacharya (1974), David and Galambos (1974), David and Nagaraja (1998), Wang (2008). Denote by $F_{i:n}(x, y)$ the joint distribution of order statistic $X_{i:n}$ and its concomitant $Y_{[i:n]}$.

Let X_1, X_2, \dots, X_n be a univariate sample with cdf $F(x)$ and $F_{i:n}(x) = P\{X_{i:n} \leq x\}$, where $X_{i:n}$ is the i th order statistic of this sample. Recently, Bairamov (2011) considered mixtures of distribution functions of order statistics

$$K_n(x) := \sum_{i=1}^n p_i F_{i:n}(x) \text{ and } H_n(x) := \sum_{i=1}^n p_i F_{n-i+1:n}(x) \text{ and using inequalities of}$$

majorization theory showed that for a particular choice of p_i 's, $H_n(x) \leq F(x) \leq K_n(x)$ for all $x \in \mathbb{R}$ and the L_2 distance between $H_n(x)$ and $K_n(x)$ can be made sufficiently small. In other words there exists a sequence $(p_1, p_2, \dots, p_n) = (p_1(m), p_2(m), \dots, p_n(m))$ such that for this sequence $H_n(x)$ and $K_n(x)$ converge to $F(x)$ as $m \rightarrow \infty$ with rate $o(1/x^{1+\alpha})$, $0 < \alpha < 1$ and the L_1 distance between $H_n(x)$ and $K_n(x)$ can be made as small as required.

In this paper we extend the results presented in Bairamov (2011) to the bivariate case. We consider mixtures of joint distribution functions of order statistics and concomitants $K_n(x, y) := \sum_{i=1}^n p_i F_{i:n}(x, y)$ and $H_n(x, y) := \sum_{i=1}^n p_i F_{n-i+1:n}(x, y)$. Using majorization inequalities it is shown that for a particular sequence $(p_1, p_2, \dots, p_n) = (p_1(m), p_2(m), \dots, p_n(m))$, $m = 1, 2, \dots$, it is true that $H_n(x, y) \leq F(x, y) \leq K_n(x, y)$ for all $(x, y) \in \mathbb{R}$ and the distance between $H_n(x, y)$ and $K_n(x, y)$ goes to zero as $m \rightarrow \infty$.

2 Auxiliary Results

Let $\mathbf{a} = (a_1, a_2, \dots, a_n) \in \mathbb{R}^n$, $\mathbf{b} = (b_1, b_2, \dots, b_n) \in \mathbb{R}^n$ and $a_{[1]} \geq a_{[2]} \geq \dots \geq a_{[n]}$ denote the components of \mathbf{a} in decreasing order. The vector \mathbf{a} is said to be majorized by the vector \mathbf{b} and denoted by $\mathbf{a} \prec \mathbf{b}$, if

$$\sum_{i=1}^k a_{[i]} \leq \sum_{i=1}^k b_{[i]} \text{ for } k = 1, 2, \dots, n-1$$

and

$$\sum_{i=1}^n a_{[i]} = \sum_{i=1}^n b_{[i]}.$$

The details of the theory of majorization can be found in Marshall et al. (2011). The following two theorems are important for our study.

Proposition 1 Denote $D = \{(x_1, x_2, \dots, x_n) : x_1 \geq x_2 \geq \dots \geq x_n\}$, $\mathbf{a} = (a_1, a_2, \dots, a_n)$, $\mathbf{b} = (b_1, b_2, \dots, b_n)$. The inequality

$$\sum_{i=1}^n a_i x_i \leq \sum_{i=1}^n b_i x_i$$

holds for all $(x_1, x_2, \dots, x_n) \in D$ if and only if $\mathbf{a} \prec \mathbf{b}$ in D . (Marshall et al. 2011, page 160).

Proposition 2 The inequality

$$\sum_{i=1}^n a_i x_i \leq \sum_{i=1}^n b_i x_i$$

holds whenever $x_1 \leq x_2 \leq \dots \leq x_n$ if and only if

$$\sum_{i=1}^k a_i \geq \sum_{i=1}^k b_i, \quad k = 1, 2, \dots, n-1$$

$$\sum_{i=1}^n a_i = \sum_{i=1}^n b_i.$$

(Marshall et al. 2011, page 639).

3 Main results in bivariate case

Let (X, Y) be absolutely continuous random vector with joint cdf $F(x, y)$ and probability density function (pdf) $f(x, y)$. Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be independent copies of (X, Y) . Let $X_{r:n}$ be the r th order statistic and $Y_{[r:n]}$ be its concomitant, i.e. $Y_{[r:n]} = Y_i$ iff $X_{r:n} = X_i$. The joint distribution of $X_{r:n}$ and $Y_{[r:n]}$ can be easily derived and it is

$$F_{r:n}(x, y) = B_n \int_{-\infty}^x F_X^{r-1}(u)(1 - F_X(u))^{n-r} F(du, y) du$$

$$= B_n \int_{-\infty}^x F_X^{r-1}(u)(1 - F_X(u))^{n-r} \left(\int_{-\infty}^y f(u, v) dv \right) du,$$

where

$$F(du, y) = \frac{\partial}{\partial u} F(u, y) = \int_{-\infty}^y f(u, v) dv, \quad \text{and} \quad B_n = n \binom{n-1}{r-1}.$$

It is easy to check that

$$\frac{1}{n} \sum_{r=1}^n F_{r:n}(x, y) = F(x, y). \quad (1)$$

Lemma 1 $F_{r+1:n}(x, y) \leq F_{r:n}(x, y)$, $r = 1, 2, \dots, n-1$, for all $(x, y) \in \mathbb{R}^2$.

Proof. We have

$$\begin{aligned}
& F_{r+1:n}(x, y) - F_{r:n}(x, y) \\
&= n \binom{n-1}{r} \int_{-\infty}^x F_X^r(u) (1 - F_X(u))^{n-r-1} F(du, y) du \\
&\quad - n \binom{n-1}{r-1} \int_{-\infty}^x F_X^{r-1}(u) (1 - F_X(u))^{n-r} F(du, y) du \\
&= \int_{-\infty}^x \left[n \binom{n-1}{r} F_X^r(u) (1 - F_X(u))^{n-r-1} - \right. \\
&\quad \left. n \binom{n-1}{r-1} F_X^{r-1}(u) (1 - F_X(u))^{n-r} \right] F(du, y) du \\
&= \int_0^{F_X(x)} \left[n \binom{n-1}{r} t^r (1-t)^{n-r-1} \right. \\
&\quad \left. - n \binom{n-1}{r-1} t^{r-1} (1-t)^{n-r} \right] F(dF_X^{-1}(t), y) dt, \tag{2}
\end{aligned}$$

where

$$F(dF_X^{-1}(t), y) = \int_{-\infty}^y f(F_X^{-1}(t), v) dv.$$

Since

$$h(t) := n \binom{n-1}{r} t^r (1-t)^{n-r-1} - n \binom{n-1}{r-1} t^{r-1} (1-t)^{n-r} \text{ and } g(t) := F(dF_X^{-1}(t), y)$$

are both bounded integrable functions in $t \in [0, F_X^{-1}(x)]$, for all $x, y \in \mathbb{R}$ and $g(x)$ is one sided function in this interval, then by the first mean value theorem for integral (see Gradshteyn and Ryzhik (2007), 12.1, page 1053)

$$\int_0^{F_X(x)} h(t) g(t) dt = g(\xi) \int_0^{F_X(x)} h(t) dt,$$

where $0 \leq \xi_x \leq F_X(x)$, $x \in \mathbb{R}$, $g(\xi_x) \geq 0$. The last equality together with (2) leads to

$$\begin{aligned}
& F_{r+1:n}(x, y) - F_{r:n}(x, y) \\
&= g(\xi_x) \int_0^{F_X(x)} \left[n \binom{n-1}{r} t^r (1-t)^{n-r-1} - n \binom{n-1}{r-1} t^{r-1} (1-t)^{n-r} \right] dt \\
&= g(\xi_x) \left\{ \int_0^{F_X(x)} n \binom{n-1}{r} t^r (1-t)^{n-r-1} - \int_0^{F_X(x)} \frac{n!}{(r-1)!(n-r)!} t^{r-1} (1-t)^{n-r} dt \right\} \\
&= g(\xi_x) [P\{X_{r+1:n} \leq x\} - P\{X_{r:n} \leq x\}] \leq 0.
\end{aligned}$$

■

Denote

$$D_+^1 = \{(x_1, x_2, \dots, x_n) : x_i \geq 0, i = 1, 2, \dots, n; x_1 \geq x_2 \geq \dots \geq x_n, \sum_{i=1}^n x_i = 1\}.$$

Lemma 2 Let $(p_1, p_2, \dots, p_n) \in D_+^1$. Then

$$H_n(x, y) \equiv \sum_{i=1}^n p_i F_{n-i+1:n}(x, y) \leq F(x, y) \leq \sum_{i=1}^n p_i F_{i:n}(x, y) \equiv K_n(x, y) \quad (3)$$

for all $(x, y) \in \mathbb{R}^2$

and the equality holds if and only if $(p_1, p_2, \dots, p_n) = (\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$. Furthermore, if $\mathbf{p} = (p_1, p_2, \dots, p_n) \in D_+^1$, $\mathbf{q} = (q_1, q_2, \dots, q_n) \in D_+^1$ and $\mathbf{p} \prec \mathbf{q}$, then

$$\begin{aligned}
\sum_{i=1}^n q_i F_{n-i+1:n}(x, y) &\leq \sum_{i=1}^n p_i F_{n-i+1:n}(x, y) \leq F(x, y) \leq \sum_{i=1}^n p_i F_{i:n}(x, y) \quad (4) \\
&\leq \sum_{i=1}^n q_i F_{i:n}(x, y).
\end{aligned}$$

Proof. By Lemma 1 $F_{1:n}(x, y) \geq F_{2:n}(x, y) \geq \dots \geq F_{n:n}(x, y)$ for all $x \in \mathbb{R}$, and $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \prec (p_1, p_2, \dots, p_n)$, the right hand side of the inequality (3) follows from the Proposition 1 and left hand side follows from Proposition 2. ■

Theorem 3 Let $p_i(m) = \frac{m+n-i+1}{a_n(m)}$, $i = 1, 2, \dots, n$; $m \in \{0, 1, 2, \dots\}$, where $a_n(m) = nm + \frac{n(n+1)}{2}$. Then

$$\begin{aligned}
H_n^{(m)}(x, y) &\equiv \sum_{i=1}^n p_i(m) F_{n-i+1:n}(x, y) \leq F(x, y) \quad (5) \\
&\leq \sum_{i=1}^n p_i(m) F_{i:n}(x, y) \equiv K_n^{(m)}(x, y) \text{ for all } (x, y) \in \mathbb{R}^2
\end{aligned}$$

and

$$\begin{aligned} \lim_{m \rightarrow \infty} \sum_{i=1}^n p_i(m) F_{n-i+1:n}(x, y) &= \lim_{m \rightarrow \infty} \sum_{i=1}^n p_i(m) F_{i:n}(x, y) \\ &= F(x, y) \text{ for all } (x, y) \in \mathbb{R}^2. \end{aligned} \quad (6)$$

Furthermore,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| K_n^{(m)}(x, y) - H_n^{(m)}(x, y) \right| dx dy = o\left(\frac{1}{m^{1-\alpha}}\right), \quad 0 < \alpha < 1. \quad (7)$$

Proof. Consider $p_i(m) = \frac{m+n-i+1}{a_n(m)}$, $i = 1, 2, \dots, n$; $m \in \{0, 1, 2, \dots\}$, where $a_n(m) = nm + \frac{n(n+1)}{2}$. It is clear that $p_1(m) \geq p_2(m) \geq \dots \geq p_n(m)$ and $\sum_{i=1}^n p_i(m) = 1$. Since $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \prec (p_1(m), p_2(m), \dots, p_n(m))$ then from Lemma 1 we have

$$\sum_{i=1}^n p_i(m) F_{n-i+1:n}(x, y) \leq F(x, y) \leq \sum_{i=1}^n p_i(m) F_{i:n}(x, y). \quad (8)$$

Since

$$\lim_{m \rightarrow \infty} p_i(m) = \lim_{m \rightarrow \infty} \frac{m+i}{nm + \frac{n(n+1)}{2}} = \frac{1}{n}, \quad i = 1, 2, \dots, n,$$

(6) follows. To prove (7) consider the L_1 distance between $K_n^{(m)}(x, y)$ and $H_n^{(m)}(x, y)$. We have

$$\begin{aligned} \Delta_m &\equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| K_n^{(m)}(x, y) - H_n^{(m)}(x, y) \right| dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \sum_{i=1}^n p_i(m) F_{i:n}(x, y) - \sum_{i=1}^n p_i(m) F_{n-i+1:n}(x, y) \right| dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \sum_{i=1}^n p_i(m) F_{i:n}(x, y) - F(x, y) + F(x, y) - \sum_{i=1}^n p_i(m) F_{n-i+1:n}(x, y) \right| dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left| \sum_{i=1}^n \left(p_i(m) - \frac{1}{n} \right) F_{i:n}(x, y) + \left(\frac{1}{n} - p_i(m) \right) F_{n-i+1:n}(x, y) \right| dx dy \\ &\leq \sum_{i=1}^n \left| p_i(m) - \frac{1}{n} \right| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F_{i:n}(x, y) - F_{n-i+1:n}(x, y)| dx dy \\ &\leq (p_1(m) - \frac{1}{n}) c_n = \frac{\frac{1}{m} \frac{n(n-1)}{2}}{n^2 + \frac{n^2(n+1)}{2} \frac{1}{m}} c_n, \end{aligned}$$

where $c_n = \sum_{i=1}^n \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |F_{i:n}(x, y) - F_{n-i+1:n}(x, y)| dx dy$. ■

References

- [1] Arnold, B., Balakrishnan, N. and Nagaraja, H.N. (1992) *A First Course in Order Statistics*. John Wiley & Sons, Inc.
- [2] Bairamov, I. (20011) Majorization bounds for distribution function. arXiv:1109.0141v1 [math.ST]
- [3] David, H. A. (1973) Concomitants of Order Statistics. *Bulletin of International Statistical Institute*, 45, 295–300.
- [4] David, H. A. and Galambos, J. (1974) The Asymptotic Theory of Concomitants of Order Statistics. *Journal of Applied Probability*, 11, 762–770.
- [5] David, H. (1981) *Order Statistics*. Second Edition, Wiley, New York.
- [6] David, H.A. and Nagaraja, H.N. (2003) *Order Statistics*, Third Edition, Wiley, New York.
- [7] David, H. A. and Nagaraja, H. N. (1998) Concomitants of Order Statistics, in *Order Statistics: Theory & Methods*, eds. Balakrishnan, N. and Rao, C. R., Elsevier, Amsterdam, pp. 487–513.
- [8] Barlow, R.E. and Proschan, F. (1975) *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston, Inc.
- [9] Gradshteyn, I.S., Ryzhik, I.M. (2007) *Table of Integrals, Series, and Products*. Seventh Edition. Editor A. Jeffrey, Associate Editor D. Zwillinger.
- [10] Marshal, A. W. , Olkin, I. and Arnold, B.C. (2011) *Inequalities: Theory of Majorization and Its Applications*. Second edition. Springer.
- [11] Samaniego, F.J. (2007) *System Signatures and their Applications in Engineering Reliability*. Springer.
- [12] Wang, Ke. (2008) *On concomitants of order statistics*. Dissertation, The Ohio State University.